Assignment 1

Huzaifa Jahana (215668106)
huzaifaj@my.yorku.ca

February 4, 2021

Note: This assignment is mainly for you to review mathematical background. You have to work individually. You must use this latex template to write up your solutions. Remember to fill in your information (name, student number, email) at above. Submit a PDF file from eClass before the deadline. No late submission will be accepted. No handwriting is accepted (Note that any handwritten part will not be marked). Direct your queries to Hui Jiang (hj@eecs.yorku.ca)

-

Exercise 1

(10 marks) **Q2.3 on page 64.**

Your answers:

Since we know that: $\mathbf{x}^{\mathbf{T}}\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix} \begin{bmatrix} x_1 & x_2 & \dots & x_m \end{bmatrix} = \begin{bmatrix} x_1^2 & x_1x_2 & \dots & x_1x_m \\ x_1x_2 & x_2^2 & \dots & x_2x_m \\ \dots & \dots & \ddots & \dots \\ x_1x_m & x_2x_m & \dots & x_m^2 \end{bmatrix}$ now we have: $\mathbf{x}\mathbf{x}^{\mathbf{T}} = \begin{bmatrix} x_1 & x_2 & \dots & x_m \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix} = \sum_{i=1}^{m} \mathbf{x}_i^2 = \sum_{i=1}^{m} \mathbf{x}_i \mathbf{x}_j^T$

The same principal applies to $\sum_{i=1}^{m} \mathbf{x}_i \mathbf{y}_j^T$, thus proving that both the following matrix multiplications can be vectorized due to basic properties belonging to transpose matrix summations.

Exercise 2

(10 marks) **Q2.4 on page 64, parts a) and b).**

Your answers:

A) From the textbook we know that any two matrices, $\mathbf{A} \in \mathbb{R}^{m\times n}$ and $\mathbf{B} \in \mathbb{R}^{m\times n}$, we can verify that $\mathbf{z}^\top \mathbf{A}\mathbf{x} = \operatorname{tr}\left(\mathbf{x}\mathbf{z}^\top \mathbf{A}\right)$ is true by our property as it can be applied here resulting in both sides stating the same thing just differently as the right side is transposing an extra time.

$$\begin{aligned} &\operatorname{tr}\left(\mathbf{A}^\top \mathbf{B}\right) = \operatorname{tr}\left(\mathbf{A}\mathbf{B}^\top\right) \\ &= \operatorname{tr}\left(\mathbf{B}\mathbf{A}^\top\right) = \operatorname{tr}\left(\mathbf{B}^\top \mathbf{A}\right) \\ &= \textstyle\sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} b_{ij} \end{aligned}$$

If we follow thru and transpose the right side we get: $\mathbf{z}^\top \mathbf{A}\mathbf{x} = \mathbf{x}^\top \mathbf{A}^\top \mathbf{z}$ which is equivalent.

B) Since we know that

$$\|\mathbf{z}-\mathbf{A}\mathbf{x}\|^2 = (\mathbf{z}-\mathbf{A}\mathbf{x})^\top(\mathbf{z}-\mathbf{A}\mathbf{x}) = \left(\mathbf{z}^\top - \mathbf{x}^\top\mathbf{A}^\top\right)(\mathbf{z}-\mathbf{A}\mathbf{x})$$
$$= \mathbf{z}^\top\mathbf{z} + \mathbf{x}^\top\mathbf{A}^\top\mathbf{A}\mathbf{x} - 2\mathbf{z}^\top\mathbf{A}\mathbf{x}$$

(uncompleted)

Exercise 3

(10 marks) **Q2.7 on page 65, just show the mean and variance, not the covariance.**

Your answers:

$$E\left[X_1\right] = \int\cdots\int x_1 \frac{\Gamma\left(r_0\right)}{\prod_{i=1}^m \Gamma\left(r_i\right)} \prod_{i=1}^m x_i^{r_i-1} dx_1\cdots dx_m$$
$$= \int\cdots\int \frac{\Gamma\left(r_0\right)}{\prod_{i=1}^m \Gamma\left(r_i\right)} x_1 x_1^{r_1-1} \prod_{i=2}^{m-1} x_i^{r_i-1}\left(1-\sum_{i=1}^{m-1} x_i\right)^{r_m-1} dx_1\cdots dx_{m-1}$$
$$= \frac{\Gamma\left(r_0\right)}{\Gamma\left(r_1\right)\prod_{i=2}^m \Gamma\left(r_i\right)} \frac{\Gamma\left(r_1+1\right)\prod_{i=2}^m \Gamma\left(r_i\right)}{\Gamma\left(r_0+1\right)}$$
$$= \frac{\Gamma\left(r_0\right)}{\Gamma\left(r_0+1\right)} \frac{\Gamma\left(r_1+1\right)}{\Gamma\left(r_1\right)}$$
$$= \frac{r_1}{r_0}$$

Thus $E[X_i] = \frac{r_i}{r_0}$ *proving the mean.*

$$E\left[X_i^2\right] = \frac{\Gamma\left(r_0\right)}{\Gamma\left(r_0+2\right)} \frac{\Gamma\left(r_i+2\right)}{\Gamma\left(r_i\right)}$$
$$= \frac{\left(r_i+1\right) r_i}{\left(r_0+1\right) r_0}$$
$$\operatorname{var}\left(X_i\right) = E\left[X_i^2\right] - E\left[X_i\right]^2$$
$$= \frac{\left(r_i+1\right) r_i}{\left(r_0+1\right) r_0} - \left(\frac{r_i}{r_0}\right)^2$$
$$= \frac{r_i\left(r_0-r_i\right)}{r_0^2\left(r_0+1\right)}$$

Thus proving the variance.

Exercise 4

(10 marks) **Q2.9 on page 65.**

Your answers: For Y = Ax + B:
$$\mathrm{E}(\mathbf{Y}) = A\mathrm{E}(\mathbf{X}) + \mathbf{b}$$
$$\mathrm{Var}(\mathbf{Y}) = A\,\mathrm{Cov}(\mathbf{X})A^T$$

Based on characteristic function of distributions p(y) exists. On the premise that:

$\phi_x(t) = \exp\left(it^\top\mu - \frac{1}{2}t^\top\Sigma t\right)$

$$\begin{aligned}
\phi_y(t) &= E\left[\exp\left(it^\top(Ax+b)\right)\right] \\
&= E\left[\exp\left(it^\top b\right)\exp\left(it^\top Ax\right)\right] \\
&= \exp\left(it^\top b\right) E\left[\exp\left(i\left(A^\top t\right)^\top x\right)\right] \\
&= \exp\left(it^\top b\right)\phi_x\left(A^\top t\right) \\
&= \exp\left(it^\top b\right)\exp\left(i\left(A^\top t\right)^\top \mu - \frac{1}{2}\left(A^\top t\right)^\top \Sigma\left(A^\top t\right)\right) \\
&= \exp\left(it^\top(A\mu+b) - \frac{1}{2}t^\top A\Sigma A^\top t\right)
\end{aligned}$$

Since, $\mathrm{Cov}(x,x) = \Sigma$ and $\mathbb{E}[\mathbf{x}] = \mu$ we proved the mean vector and covariance matrix as well.

Exercise 5

(10 marks) **Q2.12 on page 66.**

Your answers:

Treat this problem like 2 univariate Gaussian distributions mutual information as our question is asking the same thing in a different way.

Thus,

$$\begin{aligned}
KL(x_1,x_2) &= -\int x_1(x)\log x_2(x)dx + \int x_1(x)\log x_1(x)dx \\
&= \frac{1}{2}\log\left(2\pi\sigma_2^2\right) + \frac{\sigma_1^2 + (\mu_1-\mu_2)^2}{2\sigma_2^2} - \frac{1}{2}\left(1+\log 2\pi\sigma_1^2\right) \\
&= \log\frac{\sigma_2}{\sigma_1} + \frac{\sigma_1^2 + (\mu_1-\mu_2)^2}{2\sigma_2^2} - \frac{1}{2}
\end{aligned}$$

If $\sigma_1 = \sigma_2$ *and* $\mu_1 = \mu_2$ then KL divergence is non-negative and equals 0, which implies that x_1 *and* x_2 are independent.

Therefore, mutual information can be viewed as an information gain from the assumption that random variables are independent.. (Assumed knowledge regarding KL divergence and mutual information relating together is known from textbook page 48).

Exercise 6

(10 marks) **Q2.15 on page 66, only parts a) and b).**

Your answers:

A) $d^2 = f(x,y,z) = x^2+y^2+z^2$ for distance where $\nabla f = (2x, 2y, 2z), \nabla g = (1,1,1)$ now we need get our system of equations. $\begin{cases} x = \lambda \\ y = \lambda \\ z = \lambda \\ x^2+y^2+z^2 = 1 \end{cases}$ This results in

$x = 1/3 \quad y = 1/3 \quad z = 1/3$ as the point.